

位相的データ解析への統計的機械学習アプローチ

福水 健次

数理・推論研究系 教授

(東北大・AIMR・平岡裕章先生、草野元紀氏との共同研究)

■ 位相的データ解析 (TDA)

データの位相的・幾何的情報を抽出するための新しい方法論

キーテクノロジー = パーシステントホモロジー

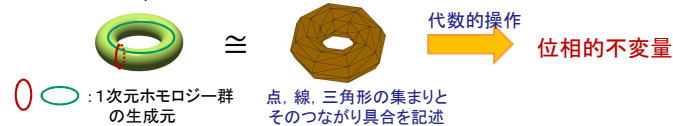
(Edelsbrunner et al 2002; Carlsson 2005)

・ 様々な応用



・ ホモロジー群 位相的不変量

図形は、三角形 (単体) の集まりで記述する \Rightarrow 代数的な扱い。



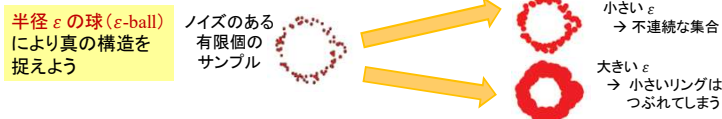
・ ホモロジー群: 位相的不変量として「穴」を群として表す。

- 0次元 = 連結成分 $H^0(X)$
- 1次元 = リング $H^1(X)$
- 2次元 = 空洞 (cavity) $H^2(X)$...

ホモロジー群の生成元: 連続的に移り合えない「穴」の代表元

・ 統計的推論における位相情報の利用

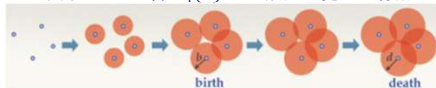
有限データからの位相の特定は容易ではない。



・ パーシステントホモロジー すべての ε を同時に考える。

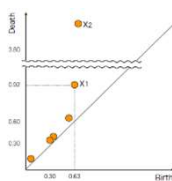
- 点集合 $X = \{x_i\}_{i=1}^m \subset \mathbb{R}^d$, $X_\varepsilon := \bigcup_{i=1}^m B_\varepsilon(x_i)$
- 位相空間の増大列 $\mathcal{X}: X_{\varepsilon_1} \subset X_{\varepsilon_2} \subset \dots \subset X_{\varepsilon_L}$ ($\varepsilon_1 < \dots < \varepsilon_L$)
- 異なるパラメータ $\varepsilon_i < \varepsilon_j$ に対し、ホモロジー生成元の関係づけが可能 (新たに発生, 継続, 消滅)。(厳密な定義はCarlsson 2009; 平岡2013)
- 各生成元の発生と消滅時刻が定まる。

1次ホモロジー群 $H_1(X)$ の生成元の発生と消滅



・ パーシステント図 (PD, 生成, 消滅の表現)

各PH生成元の発生(b), 消滅(d)時刻を、2Dグラフ上の点 (b, d) で表したものを。



複雑な幾何的データの特徴ベクトル / 記述子として使おう!

・ Persistence Weighted Gaussian Kernel

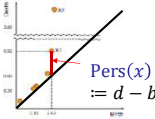
アイデア: 対角線に近い生成元はノイズの可能性が高い \rightarrow 重みを小さく

$$k_{PWG}(x, y) = w(x)w(y)\exp\left(-\frac{\|y - x\|^2}{2\sigma^2}\right)$$

- 重み関数 $w(x) := \arctan(C \text{Pers}(x)^p)$ ($C, p > 0$)

- ベクトル化により既存の統計的手法が利用可能

- 安定性定理が成立 (Kusano et al ICML2016) 点集合がHausdorff距離で微小に変化したとき、そのベクトル表現もRKHSの距離で微小にしか動かない。



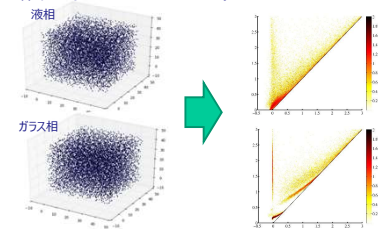
■ 応用

● シリカ(SiO_2)の液相-ガラス相転移

- 目的: 液相からガラス相に転移する温度を特定したい。
- データ: SiO_2 分子動力学 (MD) シミュレーション (Nakamura et al 2016 PNAS)



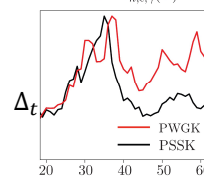
原子配置 (MDシミュレーション) パーシステント図



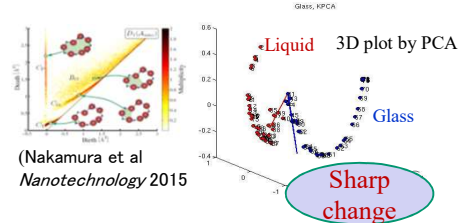
- 温度を変えて80セットの3次元原子配置データ (スナップショット) を取得
- 原子の3次元配置データから、PDを計算。

- 提案法: PDのベクトル化に対する変化点検出問題として転移点を推定

変化点検出
 $\text{KFDR}_{n, \varepsilon, \gamma}(D)$



検出された変化点 = 3100K
物理的方法: [2000K, 3500K]



● 時系列解析

- タンパク質 (1BDD) のMDシミュレーションデータ (岐阜大・一宮尚志氏提供) 60 C α . 初期状態は伸ばした状態。

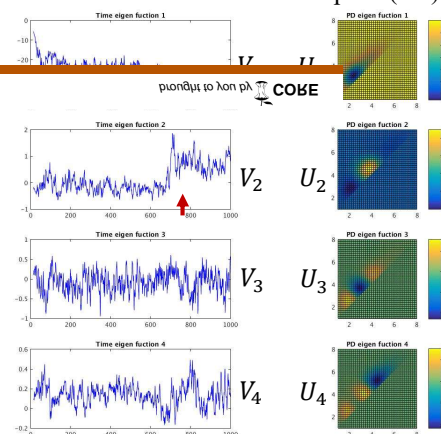
- カーネル法によるベクトル時系列をSVDによりモード分解

$$(\mu_1, \dots, \mu_t) = UTV^T$$

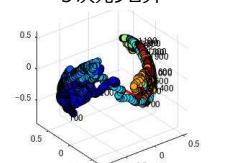
U : 空間固有ベクトル。

V : 時間固有ベクトル

time Space (PD)



カーネル主成分分析による3次元プロット



変化点 $t = 670-700$

参考文献 Kusano, G., Fukumizu, K., Hiraoka, Y. (2018) Kernel method for persistence diagrams via kernel embedding and weight factor. *Journal of Machine Learning Research*